

ゲッ歯類を用いた長期安全性試験から得られる 定量データの取り扱い

— 多群間で不等分散を示した一事例 —

小 林 克 己

(動物実験施設)

Statistical Processing Methods for Quantitative Data from Long Term Rodent Bioassays

— A Case Shown Heterogeneous in
Variance among the Groups —

Katsumi KOBAYASHI

Institute for Experimental Animals

Abstract: The statistical analysis methods of quantitative data in long term rodent bioassay, t-tests and an analysis of variance (ANOVA) including a nonparametric analysis are commonly used by the decision tree for selecting hypothesis-testing procedures. However, occasionally, no significant differences by t-tests and nonparametric data analyzed by Bartlett's test for homogeneity of variance are observed due to the outlier (abnormal) values in some groups. Those data are analyzed as the nonparametric values in case of heterogeneity in variance, and the significant differences are occasionally observed between control and treated groups, in spite of showing the same mean values in the 2 groups. Consequently, it is difficult to judge the influence of test substance on them. As the improvement methods for examining, the heterogeneous data in variance by a decision tree, outlier value which chosen by macroscopic finding is recommended to analyze by the Thompson's rejection test. And if the significant difference is observed this value will be excluded from the group as the outlier value. Those homogeneous data must be analyzed by the Student' t-test or ANOVA. An example was given to illustrate these sources of variability.

Key words: Statistical analysis, Outlier value, Rejection test, Rodent toxicity study, Heterogeneous in variance.

I はじめに

医薬品開発に必要な安全性試験, 所謂非臨床試験は厚生省のガイドライン¹⁾に従って実施されている。これら実験動物を用いた安全性試験から得られる情報データは, 飼育管理, 血液・

生化学・尿等の臨床検査および臓器重量を含めた病理学検査と極めて多く、これらデータは、最終的に統計処理を行ない薬剤の影響を検討する為の考察の手助となる。しかしながら、これら統計手法に何をどのように用いたら良いのか不明瞭な点が時折あるが、近年 t -検定系および多重範囲（比較）検定法による各々の決定樹方式^{2,3)}が常用されている。二年間の長期毒性試験では、定量データの検定項目が800から900項目（表Ⅰ）あり、この内約30%弱が薬剤の影響等によって、群間で等分散を示さないことからノンパラメトリックデータとなり順位検定によって吟味しなくてはならない⁴⁾。これらノンパラメトリックデータの中には、時々平均値がほぼ同様にもかかわらず有意差が認められることがあり、薬剤の影響があるのか、ないのか判断に迷う。

表Ⅰ ラットを用いた二年間毒性試験に於ける検定項目および検定回数

検 定 項 目	検定回数	内 容
体 重 (66+8)×2*	148	26週まで毎週、その後隔週8回増体の計算
飼 料 摂 取 料 (104+8)×2*	228	毎週測定、途中8回の総摂取量の計算
飼 料 効 率 (52+4)×2*	112	52週まで毎週、途中4回平均値の計算
血 液 学 的 検 査 (16項目×4回×2*)	128	投薬期間26, 52, 78および104週に検査
血液生化学的検査 (20項目×4回×2*)	160	同上
尿 検 査 (2項目×4回×2*)	16	同上
器官重量と体重比 (5器官×4回×2項目×2*)	80	同上
合 計	868×4	法=2,604~3,472 回

2*：雌雄

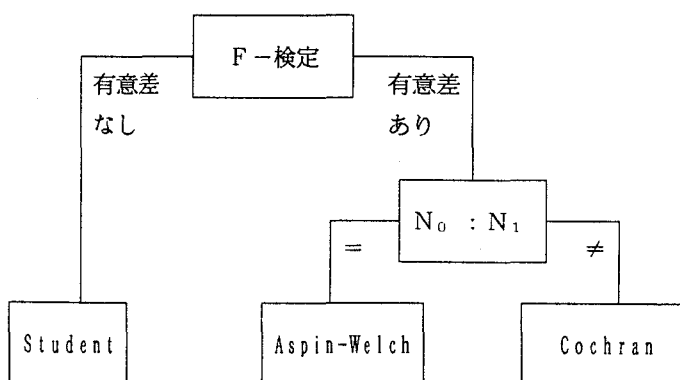
今回、これら問題点について、より適切な手法を応用するために実際の生データのどの部分に問題があるのかまた、解決する手法を検討していくつかの知見を得たので報告する。

Ⅱ 検討データ

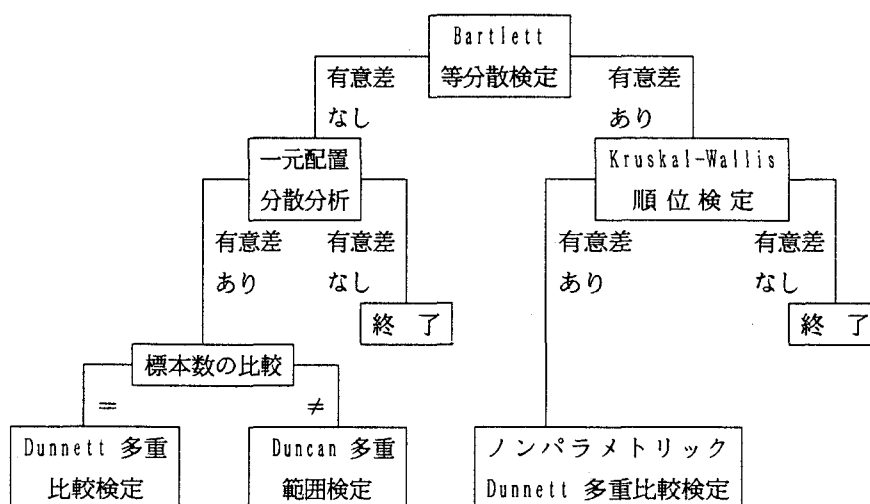
データは、F344雄ラットに対する52週後の血漿クレアチニン濃度を選んだ。試験群構成は1群20または19頭の四群（対照、低薬量、中薬量および高薬量群）で、各群の個体値は表Ⅱに示した。

Ⅲ 常用される決定樹による処理およびその結果

動物を用いた安全性試験から得られる定量データの統計手法は、一般的に決定樹を応用し t -検定法（図Ⅰ）および多重範囲（比較）検定法（図Ⅱ）を用いて対照群と各薬量群間の差を吟味している。ここの章では、多重範囲（比較）検定法を用いて上記クレアチニン濃度を検討す



図Ⅰ t-検定の決定樹



図Ⅱ 多重比較（範囲）検定の決定樹

ると。

- (1) 四群間の分散は、著しく異なっている（バートレットの等分散検定で有意差が認められた）。
- (2) 従って、クラスカル・ワリスの順位決定を実施した結果、四群間のどこかに有意差が認められた（表Ⅲ）。
- (3) 対照群と各群間との比較をノンパラメトリックダンネットの多重比較検定法で吟味した結果、対照群と高薬量群間に有意差が認められた（表Ⅳ）。

表Ⅱ F344雄ラットの投薬後52週後の
血漿クレアチニン濃度 (μ /dl)

個体番号	対照群	低薬量群	中薬量群	高薬量群
1	0.70	0.72	0.56	0.51
2	0.68	0.64	0.59	0.59
3	0.70	0.66	0.66	0.49
4	0.74	0.66	0.68	0.60
5	0.60	0.88	0.57	0.58
6	0.65	0.68	0.67	0.62
7	0.65	dead	0.70	0.51
8	0.72	0.51	0.83	0.57
9	0.63	0.65	0.86	0.60
10	0.78	0.63	0.68	2.96
11	0.67	0.79	0.60	0.56
12	0.64	0.60	0.68	0.65
13	0.63	0.69	0.57	0.71
14	0.66	0.68	0.67	0.55
15	0.88	0.62	0.53	0.54
16	0.73	0.57	0.57	0.41
17	0.57	dead	0.64	0.52
18	0.79	0.66	0.61	0.62
19	0.78	0.59	0.86	0.59
20	0.65	0.54	0.67	0.59

表Ⅲ クラスカル・ワリスの順位検定による結果

	血漿クレアチニン濃度 (μ /dl)			
	対照群	低薬量群	中薬量群	高薬量群
個 体 数	20	18	20	20
順位平均	52.2	41.2	41.9	22.5
順位総和	1043	751.5	837	449.5

四群間に、1 %以下の危険率で有意差を示す。

表Ⅳ 多重比較（範囲）検定の決定樹による結果

	個体数	クレアチニン濃度 (μ /dl)
対 照 群	20	0.69 \pm 0.07
低薬量群	18	0.65 \pm 0.09
中薬量群	20	0.66 \pm 0.10
高薬量群	20	0.69 \pm 0.54**

平均値 \pm 標準偏差。

**対照群に対して1 %以下の危険率で有意差を示す。

Ⅳ 検定結果に対する疑問点およびその理由

“対照群および高薬量の平均値は共に0.69を示したが、両群間の差は1 %以下の危険率で有意差を示した”

この理由については

- (1) クラスカル・ワリスの順位検定で実施したことによる。この検定は、パラメトリック検定のように実測値を検討するのではなく、四群すべてのデータを小さい値から順位を付け、この順位を指標に吟味する。
- (2) 個体表を見ると総個体数78の中間順位の39より大きい個体数は、対照群が15、低薬量群が10、中薬量群が11および高薬量群が3個体であったにもかかわらず、対照群と高薬量群の平均値が同一なのは、高薬量群の動物番号10の2.96が極めて大きい値を示している為で順位は小さくとも平均値は対照群と同一となる。即ち、この2.96は、たとえ0.89となっても順位は変化せず、しかし平均値は極めて小さくなり有意差が認められるならば、有意差は納得することができる。

V 有意差をどう解釈するか？

クラスカル・ワリスの順位検定およびノンパラメトリックダンネットの多重比較検定法は順位に対する有意差を示すことからこの場合、定量値については有意差を示したことになる。従って、平均値の差を吟味することができる他の検定法の応用が必要であることが示唆される。

VI 改善策として棄却検定の応用

四群の内、対照群以外の一群間に極めて大きく飛び離れた実測値が存在する場合、対照群の平均値との比較は、薬剤の影響を考察する上で繁雑となる。したがって、順位検定を行わず且つ、等分散によるt-検定または分散分析を実施したい。

4群の2.96を何らかの統計手法で外れ値（異常値）として除外し、上記の方法で実施すれば両群の差と有意水準は、パラレルに吟味できるであろう。

ゲッ菌類を用いた長期安全性試験から得られる定量データは、多種多様である。それら各測定項目では、背景値に対してどのくらいの倍率で散見されるか全く異なる。一般的に、背景値に対して飼育管理データでは、2倍以内、血液生化学検査値は、2から5倍程度また臓器重量値では、2倍から卵巣重量の1000倍程度と測定項目によって、その倍率が異なる⁵⁾。従って、背景値から各測定項目に上限（下限）値を定め疑問の値を除外することも一策である。また、棄却巾を設定しこれより逸脱した値を除外する方法⁶⁾と肉眼的に外れ値と見られる値について棄却検定を行なう方法⁶⁾があるが、なるべく個体数の減少を防ぐためにも後者の方法⁷⁾を推奨したい。

VII 棄却検定応用後の平均値の検定

トンプソンの棄却検定⁶⁾を用いて高薬量群の2.96が棄却できるか否か吟味したところ0.1%の危険率で外れ値と見なされ除外ができた。従って、高薬量群の個体数は19となる。これら四群間の比較を図IおよびIIの決定樹で再吟味した結果を表Vに示した。いずれの決定樹でも等分散が認められ対照群の平均値0.69に対して高薬量群の平均値0.57の間に有意差を示した。また、ヨンキーの傾向検定の結果は、薬量相関性を示した。

表V 高薬量群の2.96を棄却検定によって除外した後の検定結果

	個体数	クレアチニン濃度 $\mu\text{g/dl}$, (%)	スチューデント の t - 検 定	ダンカンの 多重範囲検定
対照群#	20	0.69 \pm 0.07 (100)	—	—
低薬量群	18	0.65 \pm 0.09 (94)	NS	NS
中薬量群	20	0.66 \pm 0.10 (96)	NS	NS
高薬量群	19	0.57 \pm 0.07 (83)	**	**

平均値 \pm 標準偏差。

NS対照群に対して有意差なし。

**対照群に対しての1%以下の危険率で有意差を示す。

#ノンキーの傾向検定の結果薬量相関性を示した。

VIII ま と め

ゲッ菌類を用いた長期安全性試験から得られる定量データの内、特に52週齢以降のデータには加齢による個体差の増強が加わり、また薬剤の影響による群間個体数の差異が生じ、時には著しく外れた測定値も見られることから、各群間の分散が著じるしく異なる。これらの場合、二群間の平均値の差を検定するt-検定系の決定樹では、アスピン・ウエルチおよびコ克蘭の検定法、分散分析系の決定樹では、ノンパラメトリック検定のクラスカル・ワリスの順位検定によって値を順位に変換する方法が実施されている。特に、1群に極めて著じるしく飛び離れた値がある場合、t-検定系の決定樹では有意差が認められず、一方分散分析系の決定樹では対照群間との平均値が同一にもかかわらず有意差が認められる。このような場合、薬剤の影響の有無について判断が困難となる。この原因と考えられる飛び離れた値を、外れ値として棄却検法等で除外すれば定量値自体を統計処理することができ、結果は考察に正しく反映できよう。従って、長期安全性試験から得られる定量値の統計処理では、クラスカル・ワリスの順位検定のようなノンパラメトリック検定の使用を避け、飛び離れた値を経験的に選択し、トンプソンの棄却検法等で吟味して、もし有意差が認められれば、その値を棄却し各群ともほぼ等分散化した後に、残った個体についてt-検定または分散分析を用いて検定することを推奨する。

注

本論文に掲載した統計手法の出典を下記に示す。

- (1) F-検定, スチューデントのt-検定, コ克蘭のt-検定, バートレットの等分散検定, 一元配置分散分析, ダンネットの多重比較検定, ダンカンの多重範囲検定およびクラスカル・ワリスの順位検定: Shayne Gad and Carrol S.Weil: Statistics and Experimental Design for Toxicologists. New Jersey: The Telford Press, INC., 1986.
- (2) ノンパラメトリックダンネットの多重比較検定: 佐久間昭: 薬効評価—計画と評価—II. 東京大学出

版会, 1981.

- (3) F-検定, スチューデントのt-検定, アスピン・ウエルチのt-検定, バートレットの等分散検定, 一元配置分散分析, ダンネットの多重比較検定, クラスカル・ワリスの順位検定およびヨンキーの傾向検定: 吉村 功(編): 毒性・薬効データの統計解析. サイエнтиスト社, 1987.
- (4) ヨンキーの傾向検定: Jonckheere, A.R.: A distribution-free K sample test against ordered alternatives. *Biometrika* 41:133-145, 1954.

参考文献

- 1) 厚生省薬務局審査課: 医薬品の製造(輸入)承認に必要な毒性試験のガイドラインについて, 厚生省(薬審第24号), 1989.
- 2) Shayne Gad and Carrol S.Weil: *Statistics and Experimental Design for Toxicologists*. New Jersey: The Telford Press, INC., 1986.
- 3) 吉村 功(編): 毒性・薬効データの統計解析. サイエнтиスト社, 1987.
- 4) 小林克己: 毒性試験に用いられる統計処理法の模索: 安評センター研究所報2巻: 131-141, 1988.
- 5) 小林克己: 毒性試験のための統計学: 安評センター研究所報1巻: 58-66, 1981.
- 6) 柴田寛三: 生物統計学講義, 東京農業大学, 1970.
- 7) 小林克己: 飛び離れた変量の取扱法, 医薬安全性研究会会報12号: 15-18, 1984.